

Medical Data Mining: Case of Cervical Cancer Screening

Mohamad Saraee
School of Computing, Science and
Engineering
Salford University, Salford
Manchester M5 4WT, UK

Tim Ritchings
School of Computing, Science and
Engineering
Salford University, Salford
Manchester M5 4WT, UK

Abstract

With an enormous of data stored in databases and data warehouses, data mining has been used increasingly in many fields such as retail industry, finance domain, medicine and health care, transportation, and biomedical and DNA research [1] [2]. In this paper, we are mining a large dataset of cervical cancer screening patients to try and find patterns that may have clinical significance. This work investigates different data mining software tools when applied to a large cervical cancer screening database.

1. Introduction

Cervical cancer is, potentially, one of the most preventable cancers [3]. Unlike many other cancers there is an easily detectable and normally prolonged premalignant phase. The incidence and mortality rates for cervical cancer have levelled off during the past 40 years. With early detection through Papanicolaou (Pap) test screening, cervical cancer can be prevented [4]. Minority populations and persons of low socioeconomic status, however, still have high incidence and mortality rates. The world-wide occurrence of the cancer cervix cases show that only 20% of these cases occur in the developed nations while 80% of the cases are found in the developing countries [5]. The mortality rate of this cancer can be affected by the alteration of earlier diagnosis and improved treatment in the natural history of the disease, and recent clinical research has identified that HPV is one of the main triggers for cervical cancer. The patterns discovered when mining cervical cancer screening data may support these observations, or suggest additional triggers.

2. Medical Data Mining

The data used in this study was collected within the Irisoft's Patient Management and Audit System (**COMPUSCOPE**™), which holds information related to a patient's referral and medical history, cervical examinations, pathology results, clinical management and treatments. The techniques have been used at this project such as Association Rules Mining, Clustering, and Classification are the most common methodologies in data mining [1]. After applying classification and statistical analysis to the various types of data which include numerical, integer, logical, and categorical attributes in the database, the most influential attributes were based on the frequencies that they appear. The data mining process was carried out with the focus on these attributes using the various functions of *Envisioner* a data mining tool from Neurosoft such as Relevance Analysis, Decision Tree, Classification, Clustering and Association Rules. The preliminary results will be presented in the conference.

3. References

- 1 Chen M. S., Han J., and Yu P. S., (1996), 'Data Mining: An Overview from Database Perspective', IEEE Transactions on Knowledge and Data Engineering, Vol.8, No.6, pp.866-883.
- 2 Cios K. J. (ed.), (2000), *Medical Data Mining and Knowledge Discovery*, Physica-Verlag (Springer).
- 3 BMA (1986), 'Cervical Cancer and Screening in Great Britain', Report of British Medical Association.
- 4 Posner T. and Vessey M., (1988), *Prevention of cervical cancer: the patient's view*, London: King Edward's Hospital Fund for London.
- 5 Brinton L.A., (1992), 'Epidemiology of cervical cancer-overview, the epidemiology of the cervical cancer and human papilloma virus', IACR, vol. Publication no.119.